Jeremy S Frye. Federal Scholarly Publishing: An Investigation into Practice and Public Access A Master's Paper for the M.S. in L.S degree. April, 2018. 47 pages. Advisor: Bradley M Hemminger

In the world of scholarly publishing ideals of open access have developed and thrived in a variety of fields and settings. There has been a wide variety of research published on the benefits and nature of open access, but this research has often overlooked the overlap between scholarly publishing and federal research. Federally authored works are considered works of government and considered in the public domain, but oftentimes the scholarly works of federal authors are copyrighted by private publishers and remain behind paywalls. The number of these works of government that exist behind private paywalls has not been documented and there is little research into the topic. This paper examines this phenomenon through citations gathered across major databases as well as through interviews with copy editors who work with these federally authored papers.

Headings:

Open access publishing

Scholarly publishing

Government publications

Public domain (copyright law)

# FEDERAL SCHOLARLY PUBLISHING: AN INVESTIGATION INTO PRACTICE AND PUBLIC ACCESS

by
Jeremy S. Frye

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information/Library Science.

Chapel Hill, North Carolina

April, 2018

Approved by

_____

Bradley M. Hemminger

# Table of Contents

## Section 1: Introduction

The world of scholarly publishing is a complex and often confusing world; multitudes of database and journal companies vie for the most important new articles while authors strive to publish their works in the most impactful journals that reach the largest audiences. Within this world there are numerous discourses over who is the true owner of published scholarship. Some would argue that ownership falls squarely on the copyright holder; this is oftentimes the publishing company who produces, indexes and distributes the journals. Others would argue that the authors of the papers, the people who put in the leg work and brainpower into producing the paper, should be the rightful owners of their own products. For open access advocates the answer to this question would be everyone should be the heirs to scholarly advancement. Open access scholars believe that universal access to the works of scientist and researchers the world over could bring about a new era of scholarly communication, an era that would be characterized by a free flow of information that will allow intellects from around the globe to access and build on prior research conducted by colleagues in their field (Suber, 2012) (Malamud, 2013). There have been a number of forays into the world of open access, sometimes legally through an ever-adapting publishing industry, take for example the Public Library of Science (PLoS One) or BioMed Central. Sometimes open accesses philosophy manifests in less than legal circumstances, as it did with the pirate database SciHub, a radical site that has built a large database of copyrighted scientific research that allows users to access works that should technically be restricted (Elliot, Beall, 2012).

Finally, there are many grey areas between the legal and illegal open access worlds. Take for instance the database ResearchGate which is a representative of the self-archiving practice often referred to as “green OA” (Suber, 2012) (see also arXiv) allows researchers to upload their personal draft versions of works that are copyrighted in their finalized formats. Despite the spirited debate by many open access advocates, it still remains to be demonstrated in a measurable way if open access does represent a way in which scientific communication can be improved, therefore accelerating the process of innovation and discovery that scholarly research strives to create (Antleman, 2004) (Bjork, Solomon, 2012) (Swan, 2010).

Despite the controversy surrounding many open access topics there is one area in particular that is not discussed much here in the United States, this area being the world of federal research. Federal research, like the rest of the federal government, is funded through taxpayer dollars and is therefore often thought to be in the possession of the public that paid for it. This is not always the case, oftentimes the public ends up paying for the research to be conducted, and then paying again through public institutions like libraries in order to gain access to the research. Many will be familiar with the National Institutes of Health mandate disallowing employed and funded researchers from copyrighting any products of the research conducted during the course of their official duties, be it a dataset, an official report or a scholarly journal article, it must be made openly accessible (NIH, 2008). In fact, this mandate requires all works written by the employees of or funded by NIH to eventually find their way into PubMeb Central, an open access repository of nearly 4.5 million biomedical and life science articles (PMC, 2017). But this mandate is not indicative of the whole system of federal research. There

are other institutions with similar mandates (USGS, 2017) (GPO, 2017), however, these are the exceptions rather than the rule. Furthermore, this mandate has only been around for the past decade, leaving the open status of many older federally authored papers in a questionable state.

Over the course of the past year the author engaged in a research project with a non-profit open access advocate group, in order to document exactly how many federal scholarly articles exist behind private publishing companies' database paywalls. This author's part in this project was to search for and find the majority of the body of federally written scholarly research and publish descriptions of the metadata found. Over the course of a few months around 1.2 million articles that seemed to be written or co-written by federal employees had been found and documented. The non-profit's hope for this research is to build an argument based on the fact that a number of federal mandates require works of government, works like case law, court documents or datasets, to be available to the public at minimal cost (GPO, 2017). Using these past mandates, the non-profit wants to argue that the scholarly output of federal employees counts as works of government and therefore these articles belong in the public sphere. While it is a lofty goal, the non-profit's arguments are far outside of the scope of this paper. The major use of this prior research is to demonstrate empirically that many federally authored papers are behind paywalls and to potentially understand how they got there. This research includes an audit of a small sample of results demonstrating how few federally authored papers contain disclaimers labeling them as works of government. Additionally, interviews were conducted with copyeditors who do contract work ushering scholarly papers through the publication process. The hope is that these interviews will open a

small window onto the practices of scholarly publishing and perhaps provide some answers as to why papers written by publically funded employees are inevitably sealed away in private repositories by journal and database companies.

## Section 2: Literature review

Before starting into the original research for this paper it is important to present an overview of open access and public access topics for the uninitiated. This section begins with a discussion of the history of open access. The review then transitions into the theories and laws of open and public access.

Open access has a storied history in many fields of research, as the industry of scholarly publishing began to transition into the digital world of the twenty-first century many saw the internet as a way to remove access barriers to scholarly research (Bachrach, et al. 1998). The typical barrier to access in the pre-internet age being the cost of peer-reviewing, printing and distributing, all activities typically conducted by the publisher and made up for through journal sales. Open Access scholar Peter Suber describes the effect of the internet on scholarly research as the beginning of an access revolution (Suber, 2012), surely with the free and unlimited distribution offered by the internet the barrier of cost to access journals would diminish. In the past, researchers and scholars had developed a symbiotic relationship with the journal publishing industry (Bachrach, et al. 1998). Researchers have always been motivated by a desire to have their work be disseminated and distributed to other members of their field and they were happy to provide professional organizations or independent publishers with their work in order to have it printed in journals. For scholars, payment has never been expected for the submission of their work. In fact, some of the earliest scientific journals, such as the *Philosophical Transactions of the Royal Society of London*, did not pay the authors of their articles (Guidon, 2001). This practice still continues today, and is not an issue, authors are compensated for their work in other ways. Be it salaries from their jobs or

grants from the recognition they receive, this aspect of scholarly publishing is an accepted practice. But as the price for distributing has decreased and the cost of papers continues to be near to nothing the price of access to scholarly research has only increased (Suber, 2012). In fact, the cost of library journal subscriptions has risen faster than the cost of health care in the United States (Suber, 2012). This financial strain caused by publishers on libraries has not gone undocumented (Right to Learn Coalition, 2010), according to one example the cost for serial subscriptions went up by 1300 percent over the course of thirty years at the University of California – Los Angeles (Bauerlein, M. et al, 2010). This cost is often chalked up to publishers raising prices on journal subscriptions as well as predatory bundling of collections (Strieb, K., Blixrud, J., 2013) (Green, 2017), but the truth is a little more nuanced. The information age, while not ushering in the access revolution Peter Suber hoped for, has increased the number of papers being published by a significant amount, forcing libraries to subscribe to more and more materials that are being used at much lower rates (Suber, 2012). The sheer cost of journal subscriptions is shown in an even starker light when considering the distribution of institutional wealth on an international level. While libraries in the United States have trouble affording their subscriptions to the ever-growing sea of research, libraries from less well funded university systems from around the world cannot hope to provide access to even a small majority of published materials. For example, the best funded research library in India, at the Indian Institute of Science, subscribes to 10,600 journals, on the other hand Harvard university libraries subscribe to a whopping 98,900 serials (Suber, 2012) nearly ten times as many journals. These cost issues and the dismayed hope for the

internet age led to the need for and the development of many of the foundational texts of the open access philosophy (McCabe, Snyder, 2013).

The definition of what open access truly is comes in a large part from three important works developed in the early twenty-first century. The Budapest Open Access Initiative (BOAI, 2002), the Bethesda Statement of Open Access Publishing (Brown, et al, 2003) and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (Berlin Declaration, 2003) all serve as the basis for the definition of open access in this paper and for many OA scholars worldwide (Suber, 2012). The definitions offered by these works are best summarized by Peter Suber in his 2012 work *Open Access:*

> "By "open access" to… literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited." (Suber, 2012)

To summarize this quote, open access represents a way in which scholarly research can be distributed without the typical barriers to access and restrictions of use that are part and parcel with contemporary scholarly publishing. Some barriers cannot be overcome through digital distribution alone, those without internet access cannot benefit from freely distributed digital scholarship and the philosophy of open access does not address this common technological barrier. These statements all have asserted that the authors all believe that the only point of copyright in this field of publishing is to ensure that authors are given credit for their work through proper citation practices.

In practice, there are two popular and legally endorsed practices for achieving open access. There a number of less popular and oftentimes illegal ways in which works have been brought out from behind their respective barriers, known as pirate or black open access (Green, 2017) but these methods will not be discussed in this literature review. The two main methods for achieving open access are commonly known as green and gold open access policies (Suber, 2012). Green open access refers to author self-archiving (Harnad, 2007). Self-archiving is the practice of authors providing full text copies of their works to institutional or research area repositories, this archiving typically happens before proprietary editing and formatting in order to avoid copyright violation (Suber, 2012). Green open access looks a lot like the model for public access that will be examined later in this literature review as both feature encompassing mandates for authors in certain institutions ensuring the provision of green open access copies of their works to a specific repository, be it institutional or based on the field of practice (Harnad, 2007). Green open access is widely accepted by all stakeholders in the publishing process and around 90% of the scholarly work published today has no restrictions on whether or not it can be added to a green open access repository, though sometimes an embargo period is placed on the article and the author must wait until the embargo is lifted to self-archive their work (Harnad, et al. 2013). Unfortunately, of the 90% of articles that are allowed to be deposited in open access repositories only about 20% of articles that can be self-archived have been (Harnad, et al. 2013). Many supporters of green open access argue that institutional or funder based mandates requiring authors to self-archive is the best way to grow the number of works available through green open access (Suber, 2012) (Harnad et al. 2013), but critics argue that even with mandates only about 70% of

potentially open works find their way to a repository and that green practices require fundamental changes to researchers and librarian searching strategies (Green, 2017). Green access has been seen as the preferred method for bringing works into open access as it is a cost-effective method and there have been huge successes in practice. The open access physics repository arXiv boast a nearly 100% archival rate for published physics papers (arXiv, 2018), but this may be mainly due to the culture of the field and is most likely not indicative of the practices in other research areas. Regardless of the successes or failures in this area of practice, green open access has become more popular every year with more and more institutions requiring the researchers they employee or fund to self-archive their works, see figure 2-1 for a bar graph demonstrating the number of green open access policies adopted every quarter since 2005 (ROARMAP, 2018).

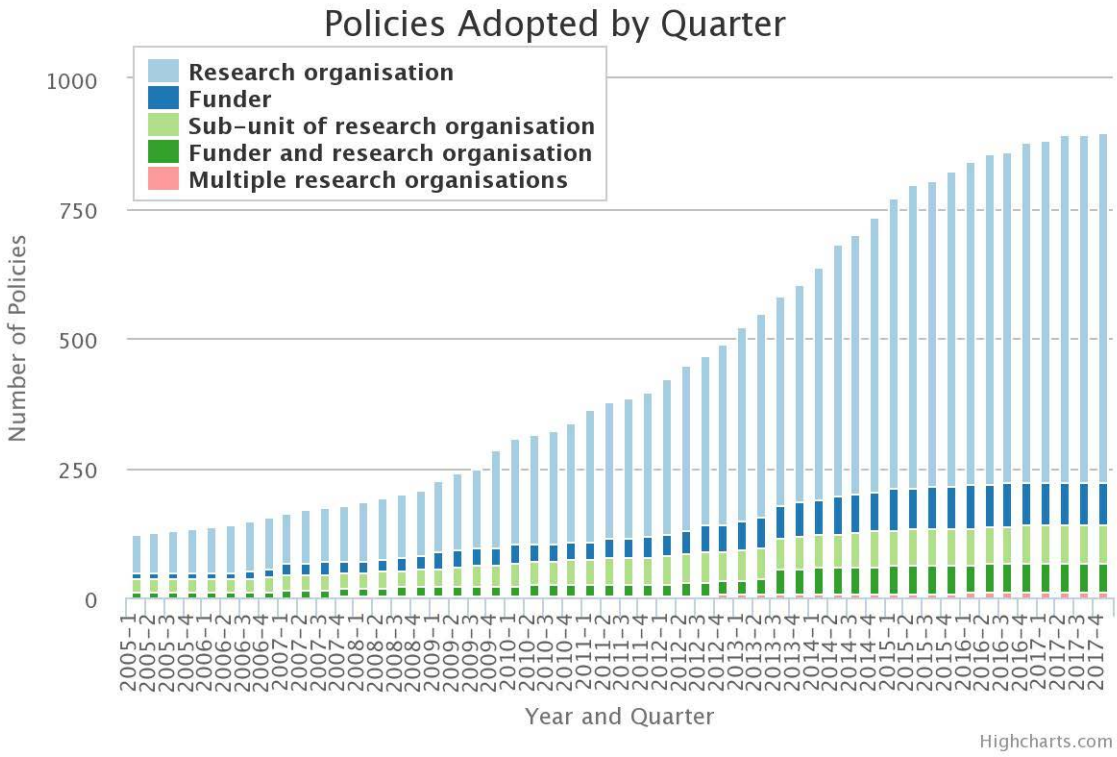


Figure 2-1: Green OA Policies Adopted by Quarter, Courtesy of ROARMAP

The other major practice to be found in the open access world is often called gold open access. This method of open access publishing revolves around authors purposefully publishing their works in open access journals, the typical practice in this area is that the author is responsible for paying for their article to be published upon submission to the open access journal (Suber, 2012). This element of pay is why this practice is referred to as gold open access and why it remains much less popular than green OA practices, only about 5% of all published materials are in gold OA journals (Harnad, et al. 2013). There have been some major successes in the gold OA field, database/journal companies like BioMed Central (BMC, 2017) and the Public Library of Science (PLoS, 2017) have thrived under the gold OA business model, using the prices incurred on authors to support expert peer-review, editorial and indexing services. While beneficial to the public by allowing scholarly works to be published OA without any embargo waiting times, the gold OA practice merely flips the burden of cost from information providers like libraries and onto institutions, funders and authors. Further criticism for gold OA comes in the form of predatory open access publishers, these publishers use unethical practices to extract publishing fees from early career authors and often do little to no work indexing or editing the papers they receive (Elliot, Beall, 2012) (DOAJ, 2018). With the cost of publishing in most major gold OA journals and the threat of predatory journals most open access scholars promote either green OA practices or the rarer practice of ‘platinum open access’ which refers to gold OA practice without the cost of publication (Elliot, Beall, 2012). Of these two major practices, green and gold OA publishing, the practice that is most reflected in the practice of the narrower field of public access in green OA publishing.

There are a number of kinds of publications related to federal scholarly publishing and each one is treated in a different way in regards to access. The major forms of information published by the United States federal government are as follows, grey literature, datasets, federally funded scholarly literature and federally authored scholarly literature. Grey literature refers to “non-conventional, fugitive, and sometimes ephemeral publications.” (Greylit.org, 2018) such as reports, administrative records, standards, federal register entries or technical documentation. As far as access to this grey material is concerned the U.S. Copyright Act of 1976 specifies that “Copyright protection under this title is not available for any work of the United States Government.” (U.S. Copyright Act of 1976), this provision has often been interpreted to mean that all of this grey literature belongs in the public domain (Bachrach, et al. 1998) and is widely accessible through web portals like Government Printing Office’s FDsys system (gpo.gov, 2018) or the National Technical Information Service (ntis.gov, 2018) or physically in one of the many federal repository libraries nationwide (fdlp.gov, 2018). Government data sets are a newer addition to the governments large output of information, but they have already been ensured to be open access by an executive order from the Obama white house (Obama, 2013). While there are ethical considerations for providing all of the data the government collects, such as the providing of government healthcare data to the public (Wade, 2007) or the potential negative effects on already marginalized populations (Raman, 2012), the release of government collected data from agencies like the Census Bureau and the EPA has become a widely accepted and lauded practice (Ubaldi, 2013) and many publically available datasets from the U.S. and abroad can be found hosted on the U.S. governments data.gov webpage (data.gov, 2018).

The two trickier federal information products when considering issues of open and public access are the two forms of scholarly output produced by government funded research, government funded papers and government written papers. There is a long legal history leading to the adoption of public access policies for government funded research. Though the amount of money provided by the government, in the form of grants to other institutions, has declined in recent years it still provided nearly 38 billion dollars to fund research in 2017 (Mervis, 2017). Until relatively recently the vast majority of these government funded papers had no official means of becoming open access, but with the advent of the National Institutes of Health open access mandate in 2008 a slow but steady pace of legislation has created channels for government funded research to be publically available (U.S. Consolidated Appropriations Act, 2008). Peter Suber described this as the first “open access mandate for a major public funding agency in the United States; it is also the first one for a public funding agency anywhere in the world that was demanded by the national legislature rather than initiated and adopted independently by the agency" (Suber, 2008). This NIH mandate required that articles written with funding through the institutes must be archived in the PubMed Central after a 12-month embargo period (Columbia, 2016). While there was a generally positive reaction to this mandate (Keener, 2010), there was also a significant backlash that led to the proposal of the bill titled the Fair Copyright in Research Works Act which would essentially make the kind of mandate the NIH had just enacted illegal (Fair Copyright in Research Works Act, 2009). This bill did not succeed fortunately and further pro-open access mandate bills were introduced in the ensuing years including the Federal Research Public Access Act (FRPAA, 2006) (FRPAA, 2010) (FRPAA, 2012) and the Fair Access to Science and

Technology Research Act (FASTR, 2013). Each of these proposals would mandate that the 11 eleven federal agencies that spend more than 100 million per year on external researchers establish a green open access policy on all funded research (Davis, 2010). While none of the proposed FRPAA or FASTR bills were enacted, other bills that would prevent federal agencies from enacting their own mandates like the Fair Copyright in Research Works Act and the Research Works Act died before being enacted as well (Fair Copyright in Research Works Act, 2009) (Research Works Act, 2011) (Cuddy, 2012). With no legal barriers preventing agencies from creating mandates similar to the NIH's and with an eventual memo from the Obama White House Office of Science and Technology Policy director Dr. John Holdren directing agencies to pursue public access policies (Holdren, 2013) a steady stream of federal mandates have manifested. So far 19 federal agencies have released public implementation guidelines (Columbia, 2016) including the Centers for Disease Control (CDC, 2015), the US Geological Survey (USGS, 2016) and National Aeronautics and Space Administration (NASA, 2014). While these mandates do a world of good for providing access to federally funded papers one can immediately tell by the titles of these statement, like the USGS's "Public Access to Results of Federally Funded Research at the U.S. Geological Survey: Scholarly Publications and Digital Data" or the CDC's "CDC Plan for Increasing Access to Scientific Publications and Digital Scientific Data Generated with CDC Funding", that this movement towards public access only involved research that is "federally funded" and not federally authored.

So, if most federally funded research is mandated to become part of the public domain after a brief embargo period, what is the access status of the scholarly articles

written by federal authors themselves? The text of the US Copyright Act of 1976 (Section 105) and the Copyright Law of 1909 (Section 7) both seem to assert that no copyright should be applied to “any work of the United States Government” (U.S. Copyright Act of 1976, Section 105) or “any reprint, in whole, or in part thereof” (Copyright Law of 1909, Section 7). These statements seem very explicit that any work coming from the US government is not subject to copyright, but this is not always the case. There is little written about the copyright status of federally authored papers, but a telling example comes from the Obama administration involving a paper written by Barack Obama himself. In 2016 Obama published a paper titled “United States Health Care Reform: Progress to Date and Next Steps” in the Journal of the American Medical Association (Obama, 2016). The paper was free to read online, but one researcher, a Michael Eisen from UC Berkeley and cofounder of PLoS, noticed something strange about the publication and began to vocally point it out: the paper had been copyrighted by JAMA (McKelvey, 2016). JAMA was quick to point out that “the journal’s copyright notice applies to the distinctive display of this JAMA article, and not the President’s work or words” (McKelvey, 2016). So, what does it mean for a copyright to be attached to a paper, but for the actual content of the paper to remain free from copyright? Is this paper publically available for public use? According to Betsy Rosenblatt, director of the Center for Intellectual Property Law at Whittier Law School, the copyrighting and occasionally the subsequent paywalling of non-copyrightable works is enacted on a regular basis because journal companies claim copyright on things like formatting and color schemes (McKelvey, 2016). So, while the authors of federally authored papers are not bound by copyright on the texts of their works, allowing for say a green OA self-archiving

approach, their works still appear to the public with an obvious copyright logo on them. While research into public understanding of copyright is outside of the scope of this paper, it is most likely fair to say that when a member of the public who does not have any experience with law sees a copyright logo on a paper they will assume that the copyright applies to the actual paper itself and not to some proprietary formatting. In fact, journal companies are not legally obligated to specify which part of a paper the copyright actually applies to, so in most cases they will not clarify that the text of a paper is free from copyright (McKelvey, 2016). Another point brought up by Rosenblatt is that, by law, journal companies are free to put whatever they like behind a paywall regardless of copyright claim (McKelvey, 2016). This point in particular begs the question, if publishers can claim copyright on a small aspect of a paper without specifying that the paper is not copyrighted and if the publisher can charge for access to the paper even if they do not own the copyright, how many works of the United States Government are currently sitting behind paywalls? Data to shine light on this very question will be presented in the analysis section of this paper.

# Section 3: Methodology

## 3.1: Existing data:

Two forms of data existing were collected for this research, the first being the citations of papers that have federal affiliations. These citations were collected over the course of a six-month fellowship for a non-profit open access group that the author took part in and represent an extensive amount of searching through large databases. The primary databases used to retrieve these citations were: ProQuest Central, Ebscohost's Academic Search Premier, Elsevier's Science Direct and Thomson Rueter's Web of Science. These databases were searched using complex Boolean strings meant to unearth every paper in the databases holdings that had even a single federal author. As each database uses slightly different syntax and has slightly different ways of representing author affiliations, unique searches had to be written for each individual database. For an example see below for a search used in the ProQuest Central database in order to discover papers written by authors with the United States Department of Agriculture:

> AF("United States Department of Agriculture" OR "USDA" OR "Farm and Foreign Agriculture Service" OR "Farm Service Agency" OR "Agricultural Stabilization and Conservation Service" OR "Farmers Home Administration" OR " Farm Security Administration" OR "Foreign Agricultural Service" OR "USDA Office of Rural Development" OR "USDA Rural Developement" OR "Rural Business-Cooperative Service" OR "Rural Housing Service" OR "Rural Utilities Service" OR "Agriculture Research Service" OR "Western Regional Research Center" OR "Southern Regional Research Center" OR "National Center for Agricultural Utilization Research" OR "Eastern Regional Research Center" OR "National Institute of Food and Agriculture" OR "Cooperative State Research, Education, and Extension Service" OR "Economic Research Service" OR "National Agricultural Statistics Service" OR "National Agricultural Library" OR "usda.gov" OR "Food and Nutrition Service" OR "FNS" OR "Center for Nutrition Policy and Promotion" OR "CNPP" OR "Food Safety and Inspection Service" OR "FSIS" OR "Natural Resources Conservation Service" OR "Soil Conservation Service" OR "United States Forest Service" OR "USFS" OR "Animal and Plant Health Inspection Service" OR "Agricultural Marketing Service" OR "Grain Inspection, Packers and Stockyards Administration" OR "Pesticide Data

> Program") NOT (Government & Official Publications AND Newspapers AND Other Sources AND Historical Newspapers AND Wire Feeds AND Working Papers)

In this search, every smaller department and agency within the larger department of the USDA was included in order to assure that no matter which way the author presented their affiliation the papers were found in the search. The searches were also refined to only retrieve scholarly journal articles, conference proceedings and academic book chapters. This refining was done in order to ensure that results represented only the scholarly output of federal authors and not works of grey literature like reports or official memos that are indexed in various databases. While the above example is representative of most searches, the Web of Science citation index allowed for easier searching using a controlled vocabulary of author affiliations. For the sake of comparison, the search used for the USDA in Web of Science looks like this “OG=(United States Department of Agriculture (USDA))”. Due to the way author affiliations are indexed in Web of Science, with all relevant institutions worked into the simple controlled vocabulary the searching was extremely simplified.

In the course of searching across these four major databases the citation information for approximately two-million scholarly journals was retrieved. These results represent searching across 82 individual government departments, agencies, branches and commissions. All major cabinet level departments were searched as well as numerous independent agencies, every major arm of the military and all federal graduate schools. National Labs, for example the Oakridge National Lab and Fermilab, were not initially searched as the work conducted at national labs is government contract work, therefore the majority of publications coming from national labs would not be considered works of

government. Eventually they were searched as well to make sure no actual government works were slipping through the cracks. Citations were sorted upon completion of the searching into 45 larger groupings with some smaller agencies being subsumed into their parent departments. The results were then de-duplicated by a colleague working with Public Resource using the Digital Object ID (DOI) as the persistent identifier across all of the citations. After duplicates were removed from the results 1,279,806 journal articles still remained, about 300,000 of which were from national labs. These results were all in the generic RIS format, which is a typical export format used by major databases and reference managers alike.

In order to verify that the results did not contain too many false positives, and that the results were in fact works of government that had been copyrighted, a convenience sample of 5797 randomly sampled DOIs were pulled from the results. A convenience sample was necessary for this process as each and every associated paper was then located using it's respective DOI and was manually inspected to verify the validity of the paper, if it had been written by a federal author or not, and to document the copyright status of the paper. Any and all disclaimer related information was taken, statements such as government disclaimers about works not representing the official stance of the federal government to statements concerning the fair use of the paper itself. Disclaimer information, author affiliations and copyright statements were harvested, along with identifying information like title, journal and publisher. These results were compiled into excel spreadsheets and organized by publisher. These spreadsheets were then unified into one encompassing sheet. The manual nature of this work necessitated the small sample size, as the process of manual examination took up a lot of time and downloading and

harvesting results using an automated bot would have violated a number of databases terms of use.

In order to assess this sample of the data a set of rules was established to determine the status of a works authorship and copyright. In order for a work to be considered written by a federal author the work had to include at least one author with a federal agency or department, not national labs, in its affiliation and there had to be no disclaimers attached clarifying that the federal author was contributing to the work outside of their official work duties. If the work met these specifications, it was labeled as having a federal employee as an author. If the work had an author from a national lab, but no clear federal authorship, the work would be labeled as having a national lab employee as an author. If the authors were neither federally affiliated or national lab employee, then the work was considered a false positive and labeled thusly. False positives typically occurred due to an international institution existing with the same name as a United States institution, for instance the Indian or Chinese Centers for Disease Control. If an agency seemed to be particularly rife with false positives, 50% or more, the search was rerun as often the false positives could be prevented with some exclusion parameters in the search. The false positives for searches that were rerun are still included in the sample.

The rules to establish if a work was copyrighted were as follows. The PDF of the paper would be inspected for any copyright or creative commons symbol; these symbols were then documented in the spreadsheet along with the institution claiming copyright if that could be determined. Further the disclaimers on the paper would be inspected, if a disclaimer said that the work was written by a federal author and therefore was a work of

government then the work was documented as a work of government, regardless of if a copyright symbol appeared on the paper. If there was no such disclaimer, then the paper was considered to be copyrighted or not considered a work of government. Once all the works had been inspected, simple COUNTIF functions were deployed in the spreadsheet in order to count the number of false positives, the number of federally authored papers, the number of national lab authored papers and the number of works that were labeled as works of government. The results of these functions and more will be discussed in the analysis section of this paper.

A second form of existing data was collected to compare the results of the above sample to the state of federal open access publishing in general. In order to obtain this data, the citation database Web of Science was utilized. When inspecting the results of a search one can use Web of Science's built in facets to see how many of the total search results are open access works. Furthermore, these facets separate results by type of open access, giving totals for gold as well as green open access. For this comparison, the ten cabinet level departments and independent agencies that had the largest research outputs were selected and searched for. The total results were documented as well as the number of open access works in both gold and green formats. There are a number of issues to consider with this method. Results were not able to be de-duplicated, meaning that any citations that appeared more than once in a single search were not able to be removed from the results during analysis. There are also papers that have coauthors from two separate agencies or departments, these were also not deduped further inflating the results for this portion of the analysis.

## 3.2: Semi-Structured Interviews

Aside from the citation data collected a few semi-structured interviews were conducted with local scholarly publishing copyeditors in the hopes of gaining an inside look into the process in which these scholarly papers are being copyrighted and published. The population interviewed were all copyeditors at a company that is contracted out by large academic publishing companies in order to edit and usher works through the publication process. These copyeditors not only edit copy, but work with the various authors and publisher based copyright offices in order to ensure that the proper forms are completed and that the copyright ends up with the rightful copyright holder. The group interviewed was selected for convenience as the author has a prior relationship with one copy editor which allowed for an easy introduction with the other employees of the company.

Three semi-structured interviews were conducted, each lasting from around twenty to thirty minutes. The participants were enlisted voluntarily and with no extrinsic motivation. The initial plan was to record the interviews and use the recordings to pull relevant quotes which would be transposed into the research, but two of the three participants requested that their responses not be recorded so the results of two of the interviews were collected through note taking during the interview process. As stated above the interviews were semi-structured, with the hopes of elucidating domain specific information from the interviewees that the author would not have thought to ask about. See the appendix for a look at the pre-planned questions. The thrust of the interviews was to ask each participant about a specific instance where they worked with federally authored papers, specifically the copyright aspect of that process.

The results of the interviews were not coded and the quotes from them are not reported verbatim due to the lack of recordings for two of the three interviews. While the sample size does seem fairly small these interviews were not intended to provide this research with rich qualitative data for analysis, rather the interviews were treated as a way to explore a small piece of the process of scholarly publication in a narrative fashion.

There are a number of considerations involved with this method. For example, one interview was conducted with a copyeditor that the author has a personal relationship with. This interview has not been included in the results of this research because there was a fear that the copyeditor's responses may be biased due to their personal relationship with the author and their knowledge of the research being conducted. The small sample size also presents an issue for analysis, with very few responses and even less recorded responses there was not enough data to code and conduct an in-depth analysis. Though as stated above the purpose of these interviews was to paint a small picture of practice rather than to analyze behaviors and attitudes towards the practice. Another issue with the smaller sample size is that all interviewees worked at the same company. There is a good chance that copyeditors at a different organization have an entirely different experience working with authors copyright needs, so the data presented can only be applied to the experiences of copyeditors working in this specific company.

# Section 4: Analysis

The analysis section will begin with a discussion of the search results and their total counts, it will then transition to an analysis of the sample taken from the search results. Afterwards the counts of open access papers taken from web of science will be compared to the results from the sample as well as the results from the search itself. Finally, the results of the interviews will be reported and analyzed. Discussion has been mixed in with results, but the final discussion on the research will take place in the conclusion section of this paper.

## 4.1: Search Results

In total 85 distinct research labs, agencies and federal departments were searched for using the major databases ProQuest, Academic Search Premier, Science Direct and Web of Science. Of the 85 only the 43 presented in table 4-1 retrieved more than ten results, many searched agencies retrieved no results, thusly only the 43 in the table have been included in the analysis. Of the over two million results initially downloaded 1,279,806 publications remained after the deduplication process. Of this 1.2 million there were 344,882 results associated with generic Department of Energy employees and 25,066 total results associated with the National Laboratories that are operated by the Department of Energy. Due to the fact that the research conducted at National Laboratories is contracted out to private and academic institutions the results retrieved have been assumed to be works of contractors and not federal employees. The same conclusion was reached with the results from the Department of Energy as a whole, since many of these papers were also authored in National Laboratories it was unsafe to assume that all 344,882 publications were written by federal employees. With the works affiliated

with the national labs and the Department of Energy subtracted from the results 909,858 potential works of federal authors remained in the search results. While it is important to note the difference between the national labs contract status and other federal agencies and departments, the publications that come out of national labs are still funded by federal taxes. Thusly the results from national labs were included in the convenience sample taken, due to the fact that the sample was taken to demonstrate how works financed by taxpayers, whether through contractors or not, are not in the public access space.

| Department/ Agency/ Lab | Publication Count |
|---|---|
| National Energy Technology Laboratory | 236 |
| Department of Labor | 253 |
| Office of the Secretary of Defense | 260 |
| Department of Defense (Excluding Graduate Schools and Military Arms) | 312 |
| Department of Education | 439 |
| Idaho National Laboratory | 506 |
| Mineral Management Service | 598 |
| Nuclear Regulatory Commission | 662 |
| National Science Foundation | 695 |
| Department of Transportation | 806 |
| Sandia National Laboratory | 884 |
| Department of Justice | 919 |
| Tennessee Valley Authority | 947 |
| Fermi National Accelerator Laboratory | 1093 |
| Army Corps of Engineers | 1099 |
| Agency of Health Research Quality | 1309 |
| United States Agency for International Development | 1352 |
| National Renewable Energy Laboratory | 1449 |
| Brookhaven National Laboratory | 1782 |
| Pacific Northwest National Laboratory | 2499 |
| Argonne National Laboratory | 2602 |
| Lawrence Livermore National Laboratory | 2680 |
| Oak Ridge National Laboratory | 3440 |

| | |
|---|---|
| Lawrence Berkeley National Laboratory | 3817 |
| Los Alamos National Laboratory | 4078 |
| Department of Defense Graduate Schools | 4237 |
| United State Legislative Branch | 4892 |
| Federal Reserve System | 6559 |
| Department of Health and Human Services | 10165 |
| Department of Veterans Affairs | 12426 |
| Food & Drug Administration | 22178 |
| Environmental Protection Agency | 24630 |
| United States Air Force | 32864 |
| Smithsonian Institute | 37492 |
| Centers for Disease Control | 45354 |
| Department of Interior | 45810 |
| United States Army | 52633 |
| United States Navy | 64965 |
| Department of Commerce | 85862 |
| National Aeronautics & Space Administration | 111850 |
| Department of Agriculture | 112753 |
| National Institutes of Health | 225537 |
| Department of Energy | 344882 |
| Total: | 1279806 |

Table 4-1: Search results counts by agency/department/laboratory

The results of the analysis of the sample are as follows. Of the 5,798 individual publications inspected from the randomized set it was found that 3,018 of those results were authored or coauthored by a federal employee. 1519 results had national lab affiliates and 45 results included both national lab and federal authors. The remaining results included 53 results for contractors outside of national labs, mostly from contractors at NASA's Jet Propulsion Research Center. There were also a few false positives in the sampled group, of which 265 were parts of searches that were rerun and therefore removed from the final results and 198 which were less clear on why they were retrieved and therefore could not have searches rerun. There were also a number of

papers that the authors affiliations were unclear based on the metadata associated with the papers, there were 340 of these publications total in the sample. Finally, the remaining results were not analyzed due to issues of access, 343 publications were not accessible through UNC's library subscriptions and 17 had incorrect DOIs so the original papers were not able to be located. The results of this analysis can be found below in Table 4-2:

| Result: | Count: |
|---|---|
| Federal Employees: | 3018 |
| National Lab Contractor | 1519 |
| National Lab & Federal Employee | 45 |
| Non-National Lab Contractor | 53 |
| False Positives: | 198 |
| False Positives w/ Search Reruns: | 265 |
| Unsure/Unclear: | 340 |
| No Access for Evaluation: | 343 |
| Incorrect DOI/ Couldn't Locate: | 17 |

Table 4-2 Result Type by Count

So, of all the results in the sample around 52% percent came out to be authored by federal employees, 26% were authored by national lab workers and 8% were false positives. The remaining percentages make up various other ways authorship presented itself, federal and national lab authors on one paper, or different issues of access that prevented analysis. While the sample size is very small and not statistically generalizable, if one were to extrapolate the results of this analysis to the total results retrieved during the search one might conclude that at least 52% or 665,499 papers found in the original search were papers written by federal authors, while 26% or 332,749 papers have national lab employees affiliated with them, and that at most 102,384 results were false positives before searches were rerun. The percent of false positives that did not have their

searches rerun was about 3% of the total results, meaning that actual number of false positives in the remaining results may be closer to 43,704 papers. While as stated earlier the sample size is too small to assume that these generalizations are accurate it is worth noting that the generalized number of 332,749 national lab authored papers is strikingly close to 344,882, which is the actual number of papers retrieved from the Department of Energy and associated labs during searching.

Beyond the basic counts of paper authorship an analysis was also run on the work of government status of the sample set. Of the 5798 sampled papers, only 124 contained a disclaimer stating that the work was authored by a federal employee and therefore was a work of government. Aside from this there were 4511 papers with no disclaimer information and 1163 papers that were not inspected either due to access issues or because they were false positives. See figure 4-1 for a graphical representation of these results. Of the 124 works with disclaimers labeling them as works of government 103 of

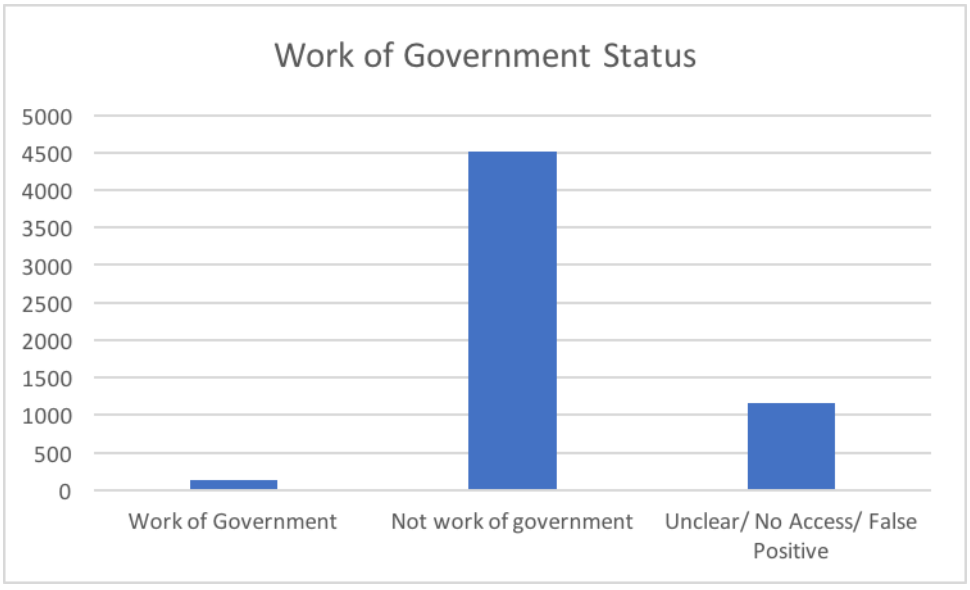


Figure 4-1 Work of Government Status of Sampled Papers

those results were papers written by federal authors, 14 of those results were written by authors at national labs, 1 was written by a non-national lab contractor and the remaining 6 were papers written by combinations of contractors, national lab employees, and federal employees. The first striking thing about these results is the tiny proportion of them that are officially listed as works of government. Of the sample inspected only 2% of the total works were publically accessible due to their nature as works of government. If national labs and false positives are excluded from analysis and only the proportion of definite federally authored papers is considered the percentage of works of government climbs to 4% of the total, but that number is still very low. It is also interesting to note that 14 of the papers labeled as works of government by publishers were works written by authors affiliated with national labs. This could indicate a number of things, potentially that these authors were federal employees and not contractors, or that certain works written by federal contractors are finding their way into the public access domain.

From the analysis of the sample a few things may be concluded. The search, while broad and encompassing national lab contractors and containing a number of false positives, did retrieve a large portion of correct hits when federal authors were being targeted. Secondly, the vast majority of papers, by federal authors as well and national lab employees, are not considered to be works of government or are not labeled thusly by publishers. Thirdly, there are papers that fall outside of this papers definition of works of government that have still been labeled as such by their publishers.

## 4.2: Web of Science Analysis

For the second portion of this analysis the results from the search and sample will be compared to the results of a number of targeted Web of Science searches, as well as to

the reported proportion of open access papers in the Web of Science searches results. The top ten federal departments and agencies with the most publications were searched and the results are presented in table 4-4. Due to the fact that Web of Science's organization index controlled vocabulary was used for the search these top ten department/agencies are different from the top ten in this papers search results. This is mainly due to the fact that all of the military arms have been included in the search for the general Department of Defense. Looking at the differences between the non-open access and the open access title counts in the table, one can see that the vast majority of papers coming from federally affiliated authors are not published in an open access or public access format. The numbers reported from these searches are also higher than the ones from the comprehensive searches conducted through multiple databases. This is the case because the results presented here have not been deduped or sampled to identify false positives.

| Department/ Agency | Not-OA | Open Access Titles: |
|---|---|---|
| Department of Energy (DOE) | 436,473 | 61,498 |
| National Institutes of Health (NIH) | 261,033 | 128,017 |
| Department of Defense (DoD) | 229,658 | 30,151 |
| Department of Agriculture (USDA) | 198,029 | 29,648 |
| National Aeronautics & Space Administration (NASA) | 115,716 | 37,989 |
| Centers for Disease Control (CDC) | 54,219 | 27,876 |
| Department of Interior (DOI) | 59,984 | 10,572 |
| Department of Veteran's Affairs (VA) | 30,602 | 16,783 |
| Food & Drug Administration (FDA) | 35,522 | 7,942 |
| Environmental Protection Agency (EPA) | 32,226 | 7,029 |

Table 4-3 Non-OA to OA Titles in Web of Science Results

Of these ten departments, the National Institutes of Health have the largest number (128,017) and the highest proportion (33%) of papers in open access. This is not surprising as the first federal agency to institute an open access mandate for papers

funded through the agency, it would not come as a shock to see papers written by NIH employees also finding their way into open access journals. The department with the smallest proportion of open access papers is the Department of Energy, with only 12% of its total publications falling in open access formats. This is to be expected, as mentioned above the Department of Energy relies heavily of contracted researchers, therefore there is less of an expectation and no legal need for these papers to be made open access. It is clear from the results of this search as well as from the analysis of the sample that the vast majority of papers written by employees of the United States federal government are not being published in a universally accessible way. In fact, only 19.7% of the papers written by federally affiliated authors indexed in Web of Science are listed as being open access. Considering that only 4% of the federally authored works in the analyzed sample were labeled as being works of government there is a significant amount of federal research that is currently behind paywalls.

Not only does Web of Science provide the data on if a paper is open access or not, one can also find data on which form of open access the publication takes, be it gold or green open access. The green open access papers have been divided further by Web of Science into the categories “Green Published” and “Green Accepted”. Green published refers to papers that are available free of charge from an online repository, while green accepted refers to papers that are not only free of charge, but are also the final peer reviewed manuscript version of the paper. There is no category in Web of Science for papers that are open access due to their federal authorship. See figure 4-2 below for a graphical breakdown of the open access format that these federal papers are published in:

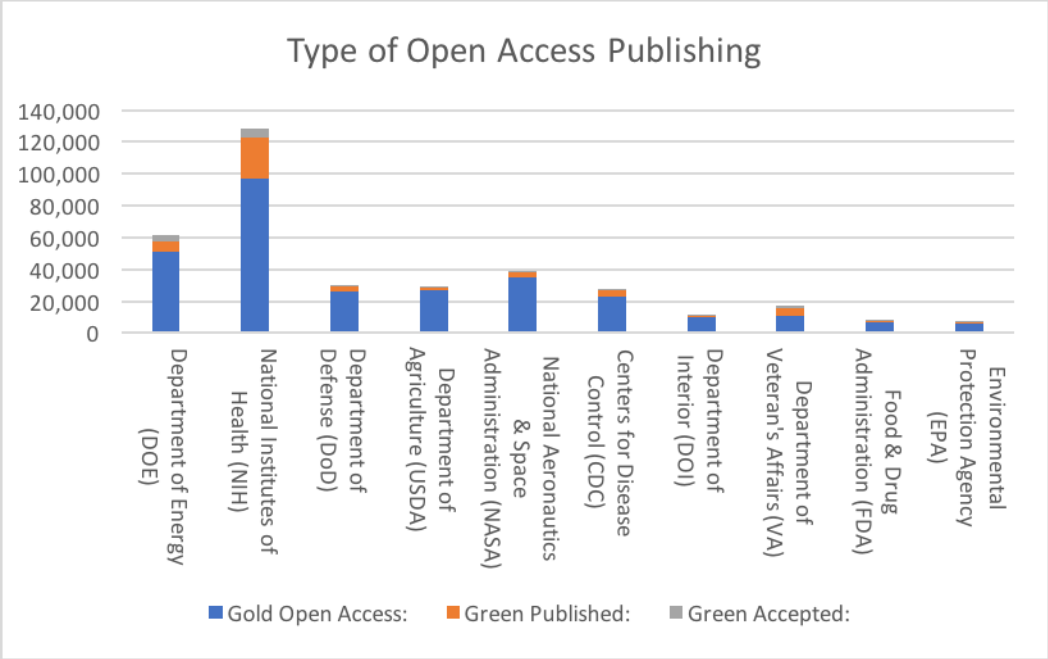


Figure 4-2 Type of OA Publishing by Department

One can immediately see from this graph that the most popular form of open access publishing for this set of federal papers is the gold open access format, of the 357,595 open access works in this analysis 292,085 (81.6%) works are published in gold open access journals. A curious tweet from the publishing house Wiley can help explain this phenomenon. A twitter user and librarian posting under the name Barbara Fister sent out a tweet with a screenshot of an article in the Wiley database (bfister, 2018). This article was clearly labeled with with a work of government disclaimer, but it was also very clearly labeled with a price tag. Fister accused Wiley of selling an article in the public domain and Wiley responded with this message, “We can confirm that articles, like this one, from US federal government employees are copyright free. However, this particular article is not open access i.e. Wiley has not received payment to publish it for free access” (Wiley, 2018). This is interesting, it could potentially explain why there is such a large proportion of government works published gold as opposed to green open

access. Perhaps federal authors are paying just to guarantee that the work is free to access, and indexed in a major database at the same time.

## 4.3 : Interviews

In total four interviews were conducted, of the four one was not included in the analysis due to a personal relationship between the participant and the interviewer that may have biased the results. Of the remaining three interviews, two of the three subjects requested that their interviews not be recorded. Therefore for two of the three interviews the data presented is merely paraphrased quotes taken at the time of the interview. Each interview lasted between fifteen and thirty minutes and participants responded to all questions posed as well as a few questions that arose during the interview. All participants had the same job title at the same organization and held similar duties. These duties consisted of editing and formatting the text of scholarly papers headed for publication, making sure that all copyright rules are followed and forms are filled out, and transmitting the papers to relevant peer reviewers, copyright offices and publishing houses. The results of these interviews will be reported by interview question when relevant, but for the most part will be written in a narrative format describing the process of publication in relation to the papers copyright.

The first question asked of participants is what topic areas are covered by the journals they work with. All participants reported working with papers on engineering, chemistry and natural sciences. While two participants who worked on freelance work outside of their regular position also stated that they worked in the topic areas of physical science and statistics. None of the participants worked with humanities, social science, law or arts based publications.

The next questions focused on the amount of federally authored scholarly papers each participant worked on during a regular work month. Each participant was first asked if they noticed the specific affiliations of authors they were working with. Two of the three participants stated that they always check affiliations first, the reasons given for this ranged from “getting to know authors” to “[knowing that] authors will have to fill out different forms depending on their affiliation”. The third participant who admitted to not initially checking author affiliations stated that they will typically notice affiliation at some point over the course of their work, especially if the author is a federal employee due to special forms these authors must fill out. The second question concerning author affiliations asked participants how often they noticed federally authored papers coming across their desks. There was a wide response to this question. One participant said they would see a federally authored paper once a month, while another stated that they only worked with federal authors once or twice a year. The third participant worked with federal authors far more, stating that about every fourth paper they worked on would have at least one federal author affiliate.

Participants were then asked about the specific actions they must take when working with a paper, federally authored or otherwise, to assign copyright. All participants expressed that the process was different depending on the specific journal and publishing company. Some journals would contract copyright work out to a separate copyright office or use an in-house copyright worker rather than assigning the tasks to copyeditors, while others would rely on the copyeditors to work with the authors to obtain copyright from them. The process of obtaining an authors copyright typically involved a form letter called a transfer of copyright form that the author must sign before

publication of their work goes through. These forms are proprietary and unique to each journal company, so the participants were not allowed or able to provide a copy of an exemplar form during the interviews. The participants described a typical form as a short document that for most authors was a basic contract that would transfer the copyright of the work from the author to the publishing company with the agreement that the publisher would print and distribute the authors work. Participants stated that this process is not always as simple as getting an authors signature and sending it off. Sometimes authors are unwilling to sign due to uncertainty of the process, and sometimes the publishing company will fail to provide additional forms for unique copyright issues. For instance when a paper has been funded by the NIH the paper must be indexed in Pubmed within a year of publication. This process requires the author to fill out another copyright form stating this and the publisher must sign their own agreement form. This process can be confusing for the copyeditors in the center of it all, as they are not trained in copyright law and are just trying to fulfill the duties of their contract. One participant described this process as often confusing for every party involved, as authors do not know the specifics of copyright, nor do copyeditors, so oftentimes the initial submission of copyright forms "can be a crapshoot as to whether [they were] submitted properly or not". Despite some confusion amongst the copyeditors all participants still described the assigning of copyright as a simple process which was mostly "determined by the publisher or journal".

This general confusion when dealing with copyright can also occur when working with federally authored papers. One participant characterized federally authored or coauthored papers as the papers that typically caused problems or confusion. All participants stated that they knew that federally authored papers needed to be treated

differently during the copyright process, but all also admitted to not knowing the specifics of why the works needed to be treated differently. The awareness of difference came from the fact that federal authors typically need to sign extra forms during the copyrighting process. One participant described the extra form for government authors in depth as a form that only applies if “all authors are government employees”, if this is the case “then the paper is in the public domain and there is no copyright to be transferred”. Participants were asked how many of work of government disclaimers they had attached to papers in the past year, responses were all fairly similar, “one or two”, “maybe two” and “I am sure of three, but I’d have to check if there were more”. When asked to characterize the papers that received disclaimers all participants revealed that the only papers that received this sort of treatment were single authored federally written papers or papers where every single co-author was employeed by the federal government. Participants were not sure what happened with a works copyright when the authorship was split between federal and other authors. The participants all described the process of adding a work of government disclaimer as similar to the typical disclaimer and copyright form process. They would receive forms from a publisher and were tasked with getting the authors signatures on all of the forms. When asked if there were any specific fields or journals that used this disclaimer more than others none of the participants could recall, but one participant said they did not believe so.

Participants were also asked about their impressions of authors knowledge when it came to copyright forms and work of government disclaimers. All participants expressed that they thought authors were the least knowledgable participants in the process dealing with their own works copyright. One quote which has been paraphrased

due to the participant declining recording sums up the views of these copyeditors: “authors just want to get their works published, and they’ll sign any form that makes the process faster”. Those interviewed were not able to recall any significant difference in author attitudes or behaviors when working with federal authors.

The final question for all interviewed asked if they could speak specifically about a federally authored paper that they could recall being difficult or at least different from the typical fair. Two participants admitted they could not remember anything out of the ordinary, but one participant provided a good example of a unique copywriting process. In this example there was a paper with majority government authors with “maybe one or two” external coauthors. The process of publication hit a snag when this copyeditor transmitted the transfer of copyright form to the paper’s authors. The government authors gave the form they had been sent to “their higher-up in their office [who read through] the form and took issue with it [because the authors] didn’t want to transfer copyright”, but given that some non-federal authors were attached to the paper this was the way the publisher had deemed the paper needed to be published. Eventually in order to satisfy the federal authors the copyeditor “split up the government and non-government authors and had them sign different forms”, with the non-government authors signing a transfer of copyright form and the federal authors signing a work of government form. The copyeditor said that this was unusual, but the publisher accepted the paper for publication. The participant was not sure what happened to the paper after it was accepted by the journal though, and could not speak as to if the work was in the public domain or not.

## Section 5 – Conclusion

So what can be discerned from the sampled data, the web of science searches and the interviews combined? From looking at solely the sample data one may conclude that there are a vast amount of federally authored papers that have not been assigned work of government disclaimers or allowed into the public domain. Some of the sample data also reveals some more confusing pieces of information, there were some national lab authored papers with works of government disclaimers even though this was thought to be something that did not happen. The data gathered from the Web of Science search could also lead one to conclude that the majority of government authored papers are behind private company paywalls, with less than 20% of papers with federal affiliates being listed as open access. The interviews also illuminate a certain amount of confusion and author based apathy when it comes to transferring the copyright of their papers. According to those interviewed most authors want their works published and are willing to do whatever they need to in order to make that happen. On top of that copyeditors working with transfers of copyright do not have the background knowledge in copyright law to know what constitutes a violation of copyright and are therefore inclined to merely transfer publisher requests regardless of legality. All of these points combined, that most papers with federal affiliated authors are not disclaimed as works of government, that only 20% of federally authored papers indexed in Web of Science are open access and that authors do not know or do not care what happens to their copyright, could lead one to conclude that there is a pervasive issue with works of government being copyrighted. But the truth seems much fuzzier than this. The interviewees all pointed out that a paper will only get special government treatment and be put in the public domain if every

single affiliated author is a government employee. This could explain away almost all of the data collected for this research. If this kind of publically accessible status is only given to papers where no authors from academia or industry assisted in any way it would not be surprising if only a few scholarly works can be considered works of government. While there is certainly room to argue against this practice, especially for papers like the final example in the interview section where the majority of authors are federal, it is out of the scope of this paper. This conclusion, that papers with any number of non-government coauthors can and should be copyrighted, does not feel like a definitive answer to the question of why there are so many papers with federal authors behind publisher paywalls. Especially given that some of the papers in the sample listed as works of government are coauthored with external authors while others are written entirely by national lab contractors, while others still have federal and academic authors despite being listed as works of government. It seems that there are no hard and fast rules defining what constitutes a work of government in the realm of scholarly publishing. This research was conducted using the definition that any paper a federal author works on is a work of the federal government, while the results indicate that this is not the case. But if some works with a private-public coauthorship can be works of government, when are they not? Furthermore if national lab contractors can author a paper that becomes a work of government does this mean that the definition of federal employee needs to be expanded as well? It seems, from this research, that in many cases there is no clear understanding or definition of what constitutes a work of government. As a result, many articles that should have been included in the public domain due to their federal authorship are relegated behind publisher paywalls.

# Bibliography

Anderson, K. (2013). Joining a CHORUS, publishers offer the OSTP a proactive, modern, and cost-saving public access solution. *The Scholarly Kitchen*, 4–6.

Anderson, R. (2004). Author disincentives and open access. *Serials Review*, *30*, 288–291. http://doi.org/10.1016/j.serrev.2004.09.001

Antelman, K. (2004). Do open-access articles have a greater research impact? *College and Research Libraries*, *65*(5), 372–382.

Bachrach, S. R., et al. (1998). Who Should Own Scientific Papers? *Science*, *281*(5382), 1459–1460.

Bailey Jr, C. W. (2005). *Open access bibliography: liberating scholarly literature with e-prints and open access journals.*

Banks, M. A., & Persily, G. L. (2010). Campus perspective on the National Institutes of Health public access policy: University of California, San Francisco, library experience. *Journal of the Medical Library Association: JMLA*, *98*(3), 256.

Bauerline, M., Gad-el-Hak, M., Grody, W., et al (2010) We must stop the avalanche of low-quality research. *Chronicle of Higher Education*

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. (2003). Retrieved from https://openaccess.mpg.de/Berlin-Declaration

Bfister. (2018, Mar. 8) “Wiley is selling an article that is in the PUBLIC DOMAIN. Says so ON THE ARTICLE ITSELF! Thankfully our ILL manager figured it out before buying it for our patrons. Scoundrels.” [Twitter Post] Retrieved from: https://twitter.com/bfister/status/971790577309233152

Bjork, B.-C., & Solomon, D. (2012). Open access versus subscription journals: a comparison of scientific impact. *BMC Medicine*, *10*(73). http://doi.org/https://doi.org/10.1186/1741-7015-10-73

Brown, P., Cabell, D., Chakravati, A., Cohen, B., Delamothe, T., Eisen, M. B., & Grivell, L. (2003). Bethesda Statement on Open Access Publishing. Retrieved from http://legacy.earlham.edu/~peters/fos/bethesda.htm

Chan, L., Cuplinkas, D., Eisen, M. B., Friend, F., & Genova, Y. (2002). Budapest Open Access Initiative.

Columbia University. (2018). Public access mandates for federally funded research. Retrieved from https://scholcomm.columbia.edu/open-access/public-access-mandates-for-federally-funded-research/

Consolidated Appropriations Act of 2008, Pub. L. No. 110-161, 121 Stat. 1844 (2008)

Centers for Disease Control (2015). *CDC PLAN FOR INCREASING ACCESS TO SCIENTIFIC PUBLICATIONS AND DIGITAL SCIENTIFIC DATA GENERATED WITH CDC FUNDING*.

Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact?. A critical review of the literature. *Journal of Infometrics*, *1*(3), 239–248.

Cuddy, C. (2012). President's Message: Open Access/Open Data. *Information Technology and Libraries (Online)*, *31*(1), 1–2.

Davis, P. (2010). The Federal Research Public Access Act (FRPAA), Science, and the Public Good. Retrieved from https://scholarlykitchen.sspnet.org/2010/04/22/frpaa-science-and-the-public-good/

Davis, P. M., Simon, D. H., & Connolly, M. J. L. (2008). Open Access Publishing, Article Downloads, and Citations: Randomised Controlled Trial. *BMj*, *337*.

Directory of Open Access Journals (2015). Principles of transparency and best practice in scholarly publishing. Retrieved from https://doaj.org/bestpractice

Doyle H., Gass, A. (2004). Who Pays for Open Access? *PLoS Biology*, *2*(4), 105.

English, R. (2006). Open Access to Federally Funded Research--The Time is Now. *Portal: Libraries and the Academy*, *6*(3), 249–252.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology*, *4*(5), 157.

Fair Access to Science and Technology Research Act of 2017, S.1701, 115th Cong. (2017)

Fair Copyright in Research Works Act, H.R.801, 111th Cong, (2009)

Federal Research Public Access Act of 2006, S.2965, 109th Cong. (2006)

Federal Research Public Access Act of 2010, S.1373, 111th Cong. (2010)

Federal Research Public Access Act of 2012, S.2096, 112th Cong. (2012)

Gargouri, Y., & Al., E. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE*, *5*(10).

GPO. (2013). GPO Strategic Plan 2014-2018. Retrieved from https://www.gpo.gov/docs/default-source/mission-vision-and-goals-pdfs/2014-2018_GPO_Strategic_Plan.pdf?sfvrsn=2

Green, T. (2017). We've failed: Pirate black open access is trumping green and gold and we must change our approach. *Learned Publishing*, *30*(4).

Harnad, S. (2010). Gold open access publishing must not be allowed to retard the progress of green open access self-archiving. *The Journal of the World Book Community*, *21*(3), 86–93.

Harnad, S. (2007). The green road to open access: A leveraged transition. In *The Culture of Periodicals from the Perspective of the Electronic Age* (pp. 99–105).

Harnad, S., Brody, T., Carr, L., Hitchcock, S., Gringa, Y., Oppenheim, C., Hilf, E. (2013). The access/impact problem and the green and gold roads to open access. *Serials Review*, *30*(4), 310–314.

Hayden, M. R. (n.d.). Open access authors express enthusiastic support and serious reservations. Retrieved from http://link.galegroup.com/apps/doc/A116077536/SCIC?u=unc_main&xid=7d1fcb56

Holdren, J. P. (2013). *Increasing access to the results of federally funded scientific research. Memorandum for the heads of executive departments and agencies. Office of Science and Technology Policy, Executive Office of the President, Washington, DC* Retrieved from www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Holdren, J. P. (2013). *Memorandum for the heads of executive departments and agencies: Increasing access to the results of federally funded scientific research*. Washington, D.C. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Howard, J. (2012). Legislation to bar public-access requirement on federal research is dead. *The Chronicle of Higher Education*, *27*.

Kaiser, J. (2015). US agencies fall in line on public access. *Science*, *348*(6231), 167.

Keener, M., Sarli, C. (2010). Public access policy support programs at libraries: A roadmap for success. *College and Research Libraries News*, *71*(10), 539–542.

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, *411*(6837), 521.

Malamud, C. (2013). On Crime and Access to Knowledge.

McKelvey, C. (2016). The copyright question behind Obama's recent publication in a medical journal. *The Daily Dot*.

Mervis, J. (2017). Data check: U.S. government share of basic research funding falls below 50%. *Science Magazine*.

NASA. (2014). *NASA Plan for Increasing Access to the Results of Scientific Research*. Retrieved from https://www.nasa.gov/sites/default/files/atoms/files/206985_2015_nasa_plan-for-web.pdf

National Institutes of Public Health. (2018). NIH Public Access Policy Details. Retrieved from https://publicaccess.nih.gov/policy.htm

Ngoie, J. K. (2014). Federal research spending and innovation in the US economy. *Journal of Policy Modeling*, *36*(3), 492–506.

Nicholas, D., & Rowlands, I. (2005). Open access publishing: The evidence from the authors. *The Journal of Academic Librarianship*, *31*(3), 179–181.

Obama, B. (2013). *Executive Order -- Making open and machine readable the new default for government information*. Washington, D.C. Retrieved from https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-

Plutchak, T. S. (2010). Searching for common ground: public access policy and the Scholarly Publishing Roundtable. *Journal of the Medical Library Association*, *98*(4), 270.

Raman, B. (2012). The rhetoric of transparency and its reality: Transparent territories, opaque power and empowerment. *The Journal of Community Informatics*, *8*(2).

Research Works Act, H.R.3699, 112th Cong. (2011)

Right to Learn Coalition (2010) *The problem: Students can't access essential research*. Retrieved from: http://old.arl.org/r2research/learn/problem/index.shtml

ROARMAP. (2018). Registry of Open Access Repository Mandates and Policies (ROARMAP).

Roth, D. L. (2008). FRPAA and NIH mandate: A blessing in disguise for scientific society publishers? *Science & Technology Libraries*, *28*(3), 247–253.

Strieb, K., Blixrud, J. (2013) The state of large-publisher bundles in 2012. *Research Library Issues*, *282*, 13-26

Suber, P. (2005). Open access, impact, and demand: Why some authors self-archive their articles. *BMJ*, *330*(7500), 1097.

Suber, P. (2008). An open access mandate for the National Institutes of Health. *Open Medicine*, *2*(2), 39–41.

Suber, P. (2012). *Open Access*. MIT Press.

Thomas, W. J., & Blackwell, L. (2010). NIH mandate one year on: How are libraries responding? *The Serials Librarian*, *58*(1–4), 257–262.

United States Geological Survey. (2016). *Public access to results of federally funded research at the U.S. Geological Survey: Scholarly publications and digital data*. United States Geological Survey, USA. Retrieved from https://www2.usgs.gov/quality_integrity/open_access/downloads/USGS-PublicAccessPlan-APPROVED-v1.03.pdf

United States Environmental Protection Agency. (2016). *Plan to increase access to results of EPA-funded scientific research*. Retrieved from https://www.epa.gov/sites/production/files/2016-12/documents/epascientificresearchtransperancyplan.pdf

U.S. Copyright Act, 17 U.S.C. §§ 101(1976, Supp. 1995)

Wade, D. (2007). Ethics of collecting and using healthcare data. *BMJ : British Medical Journal*, *334*(7608), 1330–1331.

Weitzman, J. (2004). Open access and creative common sense. *The Scientist*, *18*(9).

What is Grey Literature? (2009). Retrieved from http://greylit.org/about

WileyGlobal (2018, Mar. 13) “Thanks for raising this. We can confirm that articles, like this one, from US federal government employees are copyright free. However, this particular article is not open access i.e. Wiley has not received payment to publish it for free access” [Twitter post] Retrieved from: https://twitter.com/WileyGlobal/status/973593906087677952

Wren, J. D. (2005). Open access and openly accessible: a study of scientific publications shared via the internet. *BMJ*, *330*(7500), 1128.

Yang, Y., & Yanning, Z. (2009). The effect of open access journals on citation impact: A citation analysis of open access journals using Google Scholar. In *2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology* (pp. 278–280). IEEE. http://doi.org/10.1109/COINFO.2009.37